Using the Monte Carlo Method to Estimate Student Motivation in Scientific Computing

Isaac Caicedo-Castro[*†‡] Oswaldo Vélez-Langs[*‡] Rubby Castro-Púche[†§]
[*]Socrates Research Team
[†]Research Team: Development, Education, and Healthcare
[‡]Faculty of Engineering
[§]Faculty of Education and Human Science
University of Córdoba
Carrera 6 No. 76-103, 230002, Montería, Colombia
e-mail: {isacaic | oswaldovelez | rubycastro}@correo.unicordoba.edu.co

***Abstract*—In this study, we investigate students' motivation to learn scientific computing in the undergraduate systems engineering program at the University of Córdoba (Colombia). Scientific computing is often a challenging subject for college students; therefore, motivation plays a crucial role in succeeding in these courses. To quantify the factors that potentially impact student motivation, we conducted a survey of 117 students, relying on their perceptions. Using an F-test, we selected 15 independent variables from the original set of factors. With this dataset, we applied multidimensional linear regression to identify a function that captures the regular patterns associating motivation with its influencing factors. This prediction function quantifies a student's motivation – our target variable – based on the values of the independent variables. Using this function and the Monte Carlo method, we explored the multidimensional space of independent variables to estimate the probability that a student attains one of ten motivation levels, ranging from completely demotivated to highly motivated to learn scientific computing. Our findings indicate that students are most likely to achieve moderate motivation levels, specifically the 4th (26.86%), 5th (45.03%), and 6th (21.21%) levels. However, we also found that implementing effective policies and strategies (e.g., enhancing student satisfaction) may increase the probability of achieving higher motivation levels, particularly the 7th (36.98%) and 8th (61.04%) levels.**

Keywords-higher education; motivation; regression; Monte Carlo method.

I. Introduction

The motivation to study in college is crucial for success in courses. Demotivated students often lack the willingness, reasons, or desire to complete assignments, study for tests, and so forth. In the specific context of scientific computing courses, maintaining students' motivation can be challenging due to the difficulties associated with the mathematical concepts underlying these courses and understanding how to apply these concepts and methods to solve real-world problems.

Scientific computing courses, such as numerical methods, can be quite challenging to learn. This has prompted research aimed at predicting which students are at risk of failing these courses [1]–[4]. Indeed, these courses are difficult because they involve mathematics, programming skills, and knowledge of science for application purposes.

Studying the factors influencing the learning of mathematics has been a subject of interest in prior research, from basic educational levels [5]–[8] to higher education [9]–[13], and even doctoral levels [14]. Scientific computing, essentially an applied mathematics topic, consists of numerical methods and heuristics for solving mathematical problems in science and engineering that cannot be addressed analytically.

In Colombia, studies have examined the process of constructing knowledge among college students in the context of algebra courses within engineering curricula [10]. However, prior research has focused primarily on commitment, satisfaction, and the challenges of learning mathematics in college. As far as we know, no prior research has studied the students' motivation to learn scientific computing.

Thus, we aim to fill this gap in the literature. Our problem is to find a function that allows us to predict the student's motivation based on the variables that affect it by examining regular patterns in students' perceptions. Furthermore, we aim to explore a broader set of values for those *independent variables* than those available in the surveyed students' perceptions to calculate the probability that a student feels completely demotivated (level one) up to completely motivated (level ten). Identifying these probabilities is useful for developing policies that improve motivation among students pursuing careers that require scientific computing.

We found that it is less likely for a student to reach the highest level of motivation, while the fifth level is the most probable. Moreover, by simulating other scenarios, we found that it is possible to increase the student's motivation for learning scientific computing.

The remainder of this article is outlined as follows: In Section II, we describe how we collected and preprocessed the dataset, including its characteristics. Section III presents the regression model used to determine the functional relationship between the independent and dependent variables of this study, which forms the basis for calculating the probability of a student reaching a specific motivation level in scientific computing courses, as described in Section IV. To visualize the input variables of the regression model, we reduce the dimensionality of the input space and discuss the method used in Section V. Section VI presents and discusses our findings, while Section VII concludes the article and outlines directions for future research.

II. Collecting and Preprocessing the Dataset

To determine the aforementioned functional relationship, we collected a dataset to fit a regression model. The dataset contains independent and dependent variables. The *independent variables*, also referred to as *input variables*, were selected from *factors* that influence students' motivation for learning scientific computing. The dependent variable, also known as the target or output variable, represents the student's level of motivation.

In this study, we assumed that the factors listed in Table I influence student motivation in scientific computing. Some of these factors have been utilized in prior research [12], [13].

We conducted a survey of 117 engineering students enrolled in scientific computing courses in 2024, specifically numerical methods and nonlinear programming. The identities of the students were anonymized. Each student ranked almost all factors on a scale from one to five, with the exception of the average grade in previous mathematics courses, which is represented as a real number between zero and five (inclusive). For example, the extent to which a student felt positively about the course ranged from "not good at all" (zero) to "completely positive" (five). In contrast, the target variable was measured on a scale from zero to ten.

After collecting the dataset, we performed an F-test to select the input variables used for fitting the regression model. Input variables with a p-value less than 5×10^{-2} were selected, rejecting the null hypothesis that there is no linear relationship between the input variable and the target variable. Table I lists the input variables selected according to this criterion.

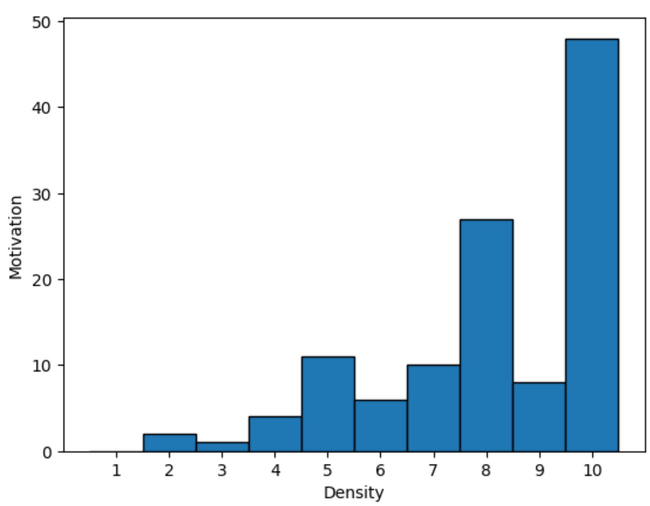

Figure 1. This histogram depicts the frequency with the students chose every level of motivation during the survey

Thus, the resulting dataset comprises 15 input variables in addition to the target variable. Each input variable forms a component of 117 vectors in a multidimensional input space, where each vector represents a student's responses to the survey, and the motivation level is recorded as the target variable. Formally, the vector $x_i \in \mathcal{X} \subset \mathbb{R}^D$ corresponds to the ith student, where the jth component x_{ij} represents the input variable associated with a specific factor. Here, D denotes the number of dimensions in the input space, i.e., $D = 15$. Additionally, the target variable $y_i \in \mathcal{Y} \subset \mathbb{R}$ represents the jth student's course motivation level.

Finally, the dataset is denoted as

$$\mathcal{D} = \{(x_i, y_i) \,|\, x_i \in \mathcal{X} \wedge y_i \in \mathcal{Y},\, i = 1, \ldots, N\} = \{(x_i, y_i)\}_{i=1}^{N},$$

where N is the number of examples in the dataset ($N = 117$), and the input vector space $\mathcal{X}$ is the Cartesian product of the domains of each input variable, i.e.,

$$\mathcal{X} = \mathbb{X}_1 \times \mathbb{X}_2 \times \cdots \times \mathbb{X}_j \times \cdots \times \mathbb{X}_D.$$

The domain of the j-th input variable is defined as

$$\mathbb{X}_j = \{a \in \mathbb{N} \,|\, 1 \leq a \leq 5\}, \quad \text{for } j = 1, \ldots, D.$$

On the other hand, the domain of the target variable is denoted as

$$\mathcal{Y} = \{a \in \mathbb{N} \,|\, 1 \leq a \leq 10\}.$$

The histogram, shown in Figure 1, illustrates that the maximum motivation level was chosen by most of the students, namely, 48 out of 117 students (see Table II).

III. Finding the Functional Relation Among Variables

The functional relationship between the target variable and the input variables is determined using *ridge regression* (a more detailed description of this method can be found in [15]). The goal is to find the function g based on the dataset $\mathcal{D}$, such that it approximates the target variable y_i given the input variables in the vector x_i, i.e., $g(x_i) \approx y_i$, where $g : \mathcal{X} \to \mathcal{Y}$. The function g is defined as a linear combination of weights and input variables as follows:

$$g(x_i) = w_0 + w_1 x_{i1} + w_2 x_{i2} + \cdots + w_D x_{iD} = w^T \hat{x}_i,$$

where $\hat{x}_i$ is an augmented vector of x_i that includes an additional component $\hat{x}_{i0} = 1$, and $\hat{x}_{ij} = x_{ij}$ for $j = 1, \ldots, D$. The components of the vector $w \in \mathbb{R}^{D+1}$ are the weights of the model. Henceforth, we shall call g the *prediction function*, as it allows us to predict a student's motivation level given the above-mentioned input variables.

For simplicity sake, the input variables are represented by the matrix $X \in \mathbb{R}^{N \times (D+1)}$, where $X_{ij} = 1$ if $j = 1$, and $X_{ij} = \hat{x}_{i,j-1}$ for $j = 2, \ldots, D+1$. Let $y \in \mathbb{R}^N$ be the vector whose components correspond to the target variable values in the dataset. This setup formalizes the problem as the following optimization problem:

$$\min_{w} f(w) = ||Xw - y||^2 + \lambda ||w||^2, \tag{1}$$

where λ is the *regularization parameter* used to prevent *overfitting* to noise in the input variables. The regularization parameter is chosen using the *elbow rule* and *10-fold cross-validation*. By setting the gradient of the objective function f with respect to w to zero and rearranging terms, the solution is obtained as:

$$w = (X^T X + \lambda I)^{-1} X^T y. \tag{2}$$

TABLE I. INPUT VARIABLES ASSOCIATED TO THE FACTORS THAT INFLUENCE THE STUDENT'S MOTIVATION IN SCIENTIFIC COMPUTING COURSES

Input Variable	*F-statistic*	*p-value*
The student's average grade in previous mathematics courses	0.43	5.16×10^{-1}
The extent to which the student has felt good about the course† $x_{i,1}$	26.17	1.27×10^{-6}
The extent to which the student has felt good about previous mathematics courses† $x_{i,2}$	24.68	2.38×10^{-6}
The extent to which the student has enjoyed the course† $x_{i,3}$	37.08	1.54×10^{-8}
The extent to which the student considers it imperative to study the course	1.08	3.02×10^{-1}
The extent to which the student considers it imperative to study mathematics courses	0.99	3.22×10^{-1}
The extent to which the student considers it wrong not to study the course	0.40	5.26×10^{-1}
The extent to which the student considers it wrong not to study mathematics courses	1.30	2.56×10^{-1}
The extent to which the student would like to recommend the course to other peers† $x_{i,4}$	37.27	1.43×10^{-8}
The extent to which the student perceives the university has up-to-date equipment† $x_{i,5}$	8.43	4.42×10^{-3}
The extent to which the course has been encouraged students to study with classmates† $x_{i,6}$	29.49	3.17×10^{-7}
The extent to which the student has been encouraged to help classmates† $x_{i,7}$	29.09	3.74×10^{-7}
The extent of the student's current engagement in participating in course lessons† $x_{i,8}$	20.43	1.51×10^{-5}
The extent of the student's current engagement in attending course lessons	2.04	1.56×10^{-1}
The extent of the student's current engagement in making an additional effort to understand the course† $x_{i,9}$	27.31	7.82×10^{-7}
The extent of the student's current focus and engagement during course lessons† $x_{i,10}$	27.59	6.96×10^{-7}
The extent to which the student has been encouraged to study the course independently† $x_{i,11}$	31.37	1.48×10^{-7}
The extent to which the student has believed the course is useful for their professional life† $x_{i,12}$	20.64	1.37×10^{-5}
The extent to which the student has considered mathematics courses useful for their professional life† $x_{i,13}$	12.94	4.75×10^{-4}
The extent to which the student has believed that they possess the ability to learn mathematics† $x_{i,14}$	3.30	7.17×10^{-2}
The extent to which the student has believed that they have the ability to solve mathematics-related problems	0.93	3.37×10^{-1}
The extent to which the student has enjoyed to solve challenging mathematics-related problems similar to those addressed in the course	15.02	1.77×10^{-4}
The extent to which the student feels their secondary school preparation is insufficient for succeeding in mathematics courses	3.67	5.79×10^{-2}
The extent to which the student believes people have innate abilities for mathematics	1.96	1.64×10^{-1}
The extent to which the student believes learning success depends on the lecturer	3.80	5.36×10^{-2}
The extent to which the student believes learning success depends on the student	1.62	2.06×10^{-1}
The extent to which the student believes hard work is key to succeeding in the course† $x_{i,15}$	4.29	4.07×10^{-2}

†The input variable is selected for regression

TABLE II. MOTIVATION LEVELS OF THE STUDENTS WHO ANSWERED THE SURVEY

Motivation Level	*Number of Students*	*Proportion of the Sample*
2	2	1.71%
3	1	0.85%
4	4	3.42%
5	11	9.40%
6	6	5.13%
7	10	8.55%
8	27	23.08%
9	8	6.84%
10	48	41.02%
Total	117	100.00%

Once the weights are computed, the function that maps the input variables to the target variable is defined. Thus, given new input variables x_{new} corresponding to a new student, the function $g(x_{\text{new}})$ calculates their respective motivation level.

IV. CALCULATING THE PROBABILITY OF EACH MOTIVATION LEVEL

In this section, we delve into the details of calculating the probability that students feel motivated at each level using the function that predicts the target variable given the input variables. To achieve this, we adopted the Monte Carlo numerical method [16].

The probability that students achieve motivation level k for learning scientific computing is defined as follows:

$$P(y_i = k) \approx P(g(x_i) = k) = \int_{\mathcal{X}} P(g(x_i) = k \mid x_i) P(x_i)\, dx_i, \tag{3}$$

where $P(x_i)$ is the probability density function of the input variables.

Assuming that each component of x_i is uniformly distributed, i.e., $x_{ij} \sim \mathcal{U}(1, 5)$ for $j = 1, \ldots, D$, the probability density function $P(x_i)$ is uniform. Therefore, Equation (3) is rewritten as:

$$P(y_i = k) \approx P(g(x_i) = k) \approx \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(g(x_i) = k), \tag{4}$$

where N is the number of vectors x_i, whose components are random numbers uniformly distributed. Moreover, $\mathbf{1}(u) = 1$ if u is true, and $\mathbf{1}(u) = 0$ otherwise. Note that N is not the size of the dataset described in Section II.

The value of N is chosen based on the standard error (SE), which is calculated as:

$$SE = \frac{\sigma}{\sqrt{N}}, \tag{5}$$

where σ is the standard deviation of the calculated probabilities. The value of N is increased iteratively until the SE decreases to a tolerable threshold.

V. Reducing the Dimensionality of the Input Space

We reduce the dimensionality of the vectors in the input space by adopting Principal Component Analysis (PCA) (see [15] for a complete description of the method). PCA transforms the original vectors into a new space with fewer dimensions, where each principal component is a linear combination of the original variables. The components are calculated to maximize variance, enabling a better understanding of the latent structure, reducing noise, and visualizing the dataset.

Let $u_1 \in \mathbb{R}^D$ be the first basis vector of the new space. The first principal component z_{i1}, corresponding to vector x_i, is calculated as:

$$z_{i1} = u_1^T x_i, \tag{6}$$

where the variance of the first principal component is defined as:

$$\mathrm{var}(z_{i1}) = u_1^T S u_1, \tag{7}$$

with $S \in \mathbb{R}^{D\times D}$ being the covariance matrix:

$$S = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T. \tag{8}$$

Here, $\bar{x}$ is the mean of the input vectors.

The optimization problem is to find u_1 that maximizes the variance of the first principal component z_{i1}, subject to the constraint that u_1 is a unit vector ($u_1^T u_1 = 1$):

$$\max_{u_1} J(u_1) = u_1^T S u_1 - \lambda_1(u_1^T u_1 - 1), \tag{9}$$

where λ_1 is a Lagrange multiplier. By setting the gradient of the objective function J to zero with respect to u_1 and simplifying, we find:

$$\lambda_1 = u_1^T S u_1 = \mathrm{var}(z_{i1}), \tag{10}$$

where λ_1 represents the variance of the first principal component.

To calculate the second principal component z_{i2}, we find the second basis vector u_2, such that $z_{i2} = u_2^T x_i$. The vector u_2 must be orthogonal to u_1 ($u_1^T u_2 = 0$) and a unit vector ($u_2^T u_2 = 1$). The optimization problem is:

$$\max_{u_2} J(u_2) = u_2^T S u_2 - \lambda_2(u_2^T u_2 - 1) - \alpha u_2^T u_1, \tag{11}$$

where λ_2 and α are Lagrange multipliers. Setting the gradient of J to zero yields:

$$\lambda_2 = u_2^T S u_2 = \mathrm{var}(z_{i2}). \tag{12}$$

This procedure generalizes to calculate all basis vectors u_j for $j = 1, \ldots, D$, producing d principal components, where $d < D$. The transformed vector $z_i \in \mathbb{R}^d$ represents the original $x_i \in \mathbb{R}^D$ in the reduced space. The vector z_i can then be used as input for the methods described in Sections III and IV.

For visualization purposes, three or fewer principal components are sufficient ($d \le 3$). Another criterion for choosing d is based on the proportion of retained variance ρ, calculated as:

$$\rho = 100 \cdot \frac{\sum_{j=1}^{d}\lambda_j}{\sum_{k=1}^{D}\lambda_k}\%. \tag{13}$$

In some cases, $\rho = 100\%$ is achieved with $d < D$ due to noise in the variables. Alternatively, a threshold for ρ (e.g., $\rho \le 97\%$) can be set depending on the application domain.

VI. Results and Discussion

The resulting prediction function g estimated through the regression model described in Section III is defined as follows:

$$\begin{aligned} g(x_i) = 0.0220 &+ 0.1678x_{i,1} + 0.1751x_{i,2} + 0.1992x_{i,3} + \ldots \\ &\cdots + 0.1989x_{i,4} + 0.1018x_{i,5} + 0.1111x_{i,6} + \ldots \\ &\cdots + 0.1592x_{i,7} + 0.1157x_{i,8} + 0.1597x_{i,9} + \ldots \\ &\cdots + 0.1557x_{i,10} + 0.1765x_{i,11} + \ldots \\ &\cdots + 0.0895x_{i,12} - 0.0049x_{i,13} + \ldots \\ &\cdots + 0.0744x_{i,14} + 0.0749x_{i,15} \end{aligned} \tag{14}$$

According to this model, the ith student's satisfaction ($x_{i,4}$) and enjoyment ($x_{i,3}$) with the scientific computing course are the variables with the highest weights; therefore, both correspond to the most influential factors in the student's motivation for the course.

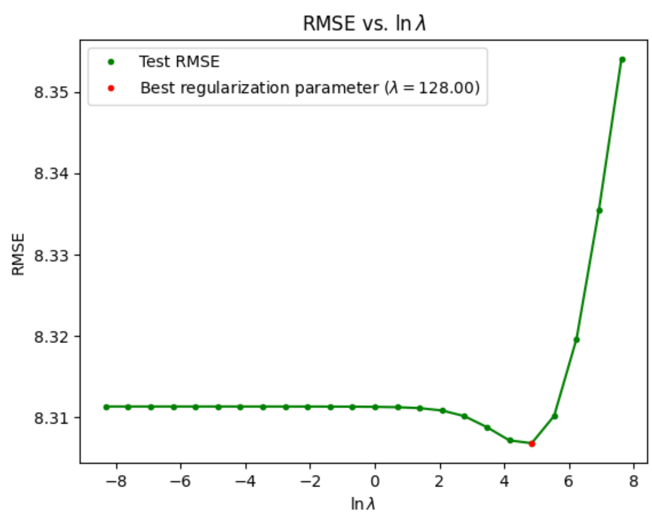

Figure 2. This chart is the elbow rule used to tune the regularization parameter of the regression model. We used values of λ ranging from 2^{-12} to 2^{12}. We adopted 10-fold cross-validation to evaluate the model with several regularization settings.

On the other hand, a negative weight for $x_{i,13}$ indicates that students who perceive mathematics courses as more useful for their careers tend to have slightly lower motivation levels in scientific computing courses. Since the absolute value of the weight is very small, the effect is weak but still worth considering. Either the students are motivated to obtain high grades in the scientific computing course, even though they consider acquiring mathematical knowledge merely a graduation requirement, or they are motivated because they can solve mathematical problems in the scientific computing course using algorithms (i.e., numerical methods or heuristics) instead of the analytical methods covered in previous classical courses (e.g., differential calculus).

The regression model achieved a coefficient of determination and a root-mean-squared-error of 0.37 and 1.62, respectively. This outcome suggests that the regression model performs better than predicting by using the mean value. The regularization parameter λ was chosen using the elbow rule as depicted in Figure 2. According to this method, the best value for regularization is $\lambda = 128$.

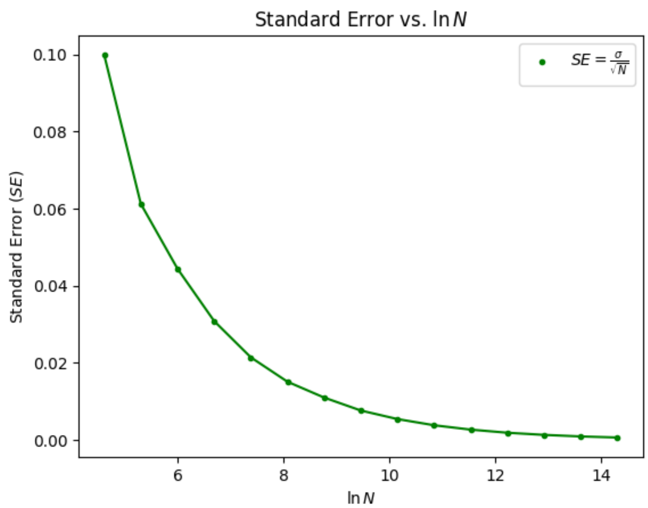

Figure 3. This chart shows how the standard error drops as the variable N is increased.

Thus, the maximum and minimum values that can be obtained from the prediction function are approximately 2 and 10, respectively. However, the most probable level according to the Monte Carlo method is 4.908 with a standard error of 6.8×10^{-4}. Figure 3 illustrates how this value was achieved as N increases. This outcome was obtained with 95% confidence (alpha = 0.05), within the interval (4.90, 4.91). However, this is not an actual level; therefore, we performed rounding to the nearest even number for halfway cases, and the probabilities of the motivation levels are shown in Table III.

It is noteworthy that the high levels of motivation have the lowest probabilities, in contrast with the motivation levels of the students in the dataset depicted in Table II. The results in Table III align with the histogram illustrated in Figure 4, where the high levels were obtained with less frequency through the Monte Carlo simulation.

TABLE III. PROBABILITY OF EVERY MOTIVATION LEVEL CALCULATED WITH THE MONTE CARLO METHOD

Level	*Probability*
1	P(y = 1.0) = 5.49×10^{-4}%
2	P(y = 2.0) = 1.34×10^{-1}%
3	P(y = 3.0) = 4.17%
4	P(y = 4.0) = 26.86%
5	P(y = 5.0) = 45.03%
6	P(y = 6.0) = 21.21%
7	P(y = 7.0) = 2.55%
8	P(y = 8.0) = 5.57×10^{-2}%
9	P(y = 9.0) = 6.10×10^{-5}%

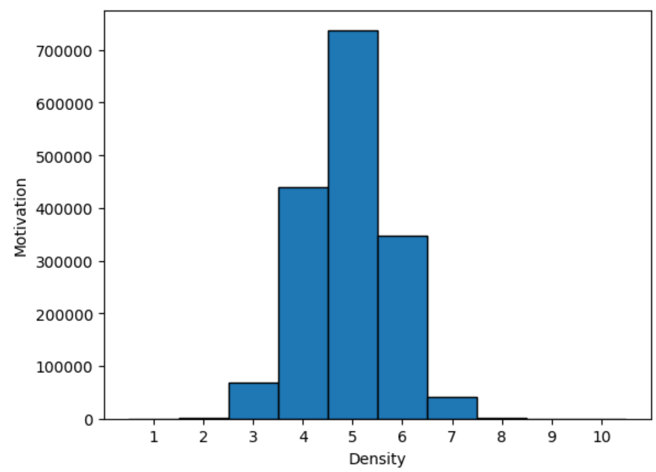

Figure 4. Histogram yielded through the Monte Carlo method. This shows the frequency with which the function g calculates each motivation level based on the random input variables.

Despite these figures, we might increase the probability of high motivation levels through policies that improve the factors associated with the input variables. We found this by simulating $P(y_i = k)$ while assuming high random values for the controllable input variables, such as $x_{ij} \sim \mathcal{U}(3,5)$ for $1 \leq j \leq 12$, and setting $x_{i,12}$ to zero to mitigate its negative influence on the prediction. Meanwhile, the variables that cannot be directly controlled were assigned random values across their entire range, i.e., $x_{ij} \sim \mathcal{U}(1,5)$ for $j = 13, 14$.

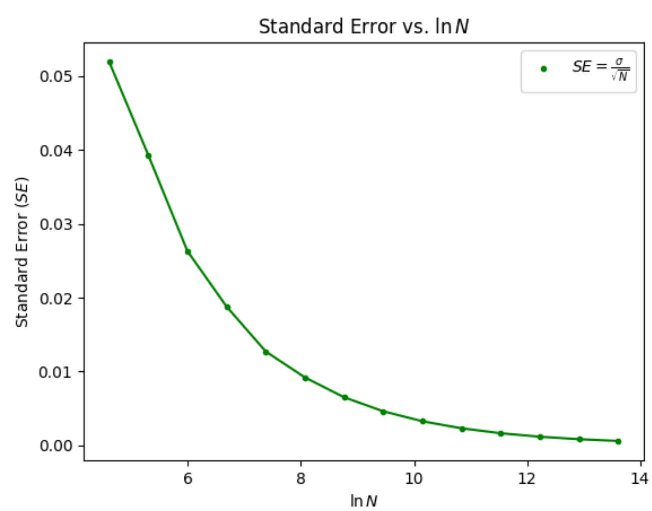

Figure 5. This chart shows how the simulation converges to the solution with a setting aiming to increase the student motivation to learn scientific computing, hence, standard error drops as the variable N is increased.

With the aforementioned simulation setting the most probable level according to the Monte Carlo method is 7.642 with a standard error of 8.1×10^{-4}. Figure 5 shows how the simulation converges as N increases. This outcome was obtained with 95% confidence (alpha = 0.05), within the interval (7.641,

Copyright (c) IARIA, 2025. ISBN: 978-1-68558-263-0

7.643). The probabilities of the motivation levels are shown in Table IV while the resulting histogram is illustrated in Figure 6.

TABLE IV. PROBABILITY OF EVERY MOTIVATION LEVEL CALCULATED WITH THE MONTE CARLO METHOD FROM THE BEST SIMULATION SETTING

Level	*Probability*
6	P(y = 6.0) = $2.5 \times 10^{-1}\%$
7	P(y = 7.0) = 36.98%
8	P(y = 8.0) = 61.03%
9	P(y = 9.0) = 1.74%

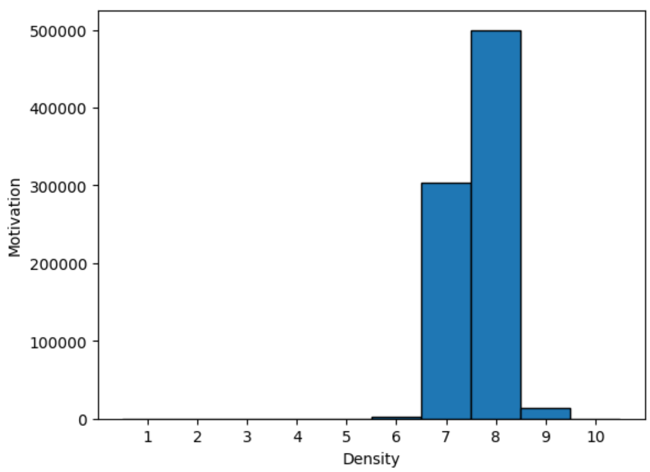

Figure 6. Histogram yielded through the Monte Carlo method using a simulation setting that aims to increase the probability of highest level of motivation for learning scientific computing

TABLE V. VARIANCE RETAINED BY THE PRINCIPAL COMPONENTS

Number of Principal Components	*Retained Variance (%)*
1	44.84%
2	55.43%
3	64.22%
4	71.16%
5	76.19%
6	80.51%
7	84.30%
8	87.74%
9	90.69%
10	92.92%
11	94.77%
12	96.50%
13	97.99%
14	99.15%
15	100.00%

The previous results reveal that designing policies to enhance satisfaction in prerequisite mathematics and scientific computing courses, providing students with up-to-date equipment, promoting teamwork, and so forth might improve students' motivation to learn scientific computing.

Additionally, regarding the visualization of the input space, using two components retained 53.43% of the variance (see Table V). To retain the complete variance, the dimensionality must be the same as the original input space or close to it. Therefore, it is pointless to use a dimensionality that cannot be visualized.

The regression model trained with a two-dimensional input space obtained through principal component analysis yields a coefficient of determination and a root-mean-squared-error of 0.33 and 1.67, respectively. Therefore, the regression model applied to the original input space outperforms the one applied to the reduced input space.

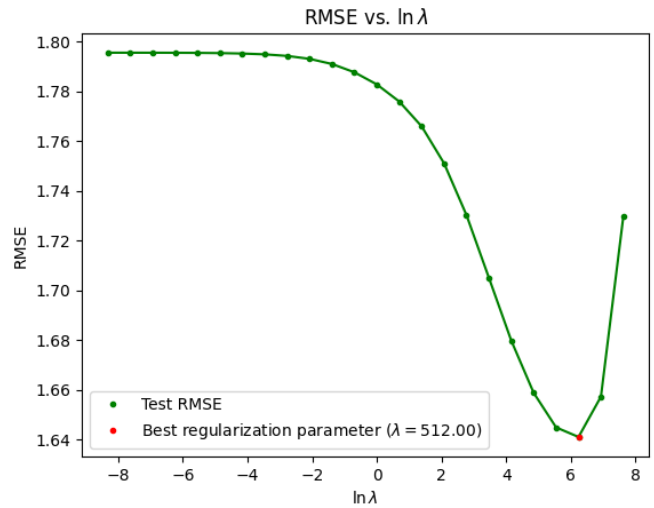

Figure 7. This chart is the elbow rule used to tune the regularization parameter of the regression model performed on a two-dimensional input space. We used values of λ ranging from 2^{-12} to 2^{12}. We utilized 10-fold cross-validation to evaluate the model with different regularization values.

Figure 7 shows how the regularization parameter was selected through the elbow rule to avoid overfitting.

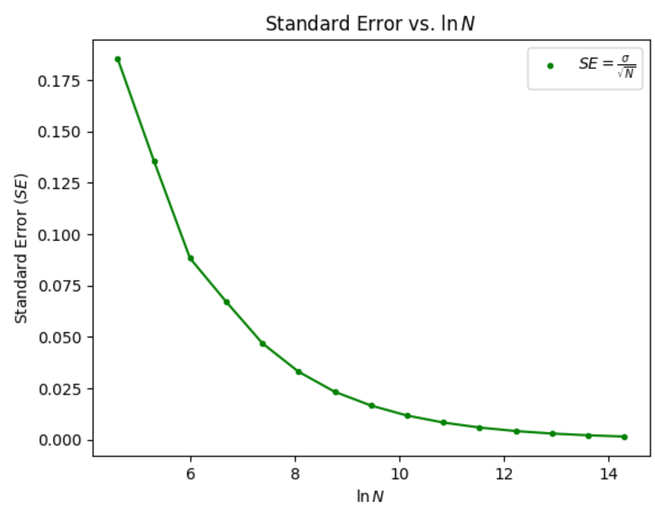

Figure 8. This chart shows how the standard error drops as the variable N is increased in the Monte Carlo simulation applied on the two-dimensional input space.

When the Monte Carlo method is performed on the regression model obtained from the new two-dimensional input space, the outcome is that the most likely level is 3.84 with a standard error of 1.46×10^{-3}. This outcome is within (3.83, 3.84) with a 95% (alpha = 0.05) confidence interval (see Figure 8)

TABLE VI. PROBABILITY OF EVERY MOTIVATION LEVEL CALCULATED WITH THE MONTE CARLO METHOD TAKING INTO ACCOUNT TWO PRINCIPAL COMPONENTS

Level	*Probability*
1	P(y = 1.0) = 13.62%
2	P(y = 2.0) = 15.61%
3	P(y = 3.0) = 15.64%
4	P(y = 4.0) = 15.59%
5	P(y = 5.0) = 15.62%
6	P(y = 6.0) = 15.59%
7	P(y = 7.0) = 8.32%

The results obtained with two-dimensional input spaces reveal that it becomes more likely that a student will reach the lower levels of motivation (see Table VI). The results align with the histogram depicted in Figure 9. In this case, the prediction function is defined as follows:

$$g(z_i) = 0.03 + 0.52z_{i1} - 0.04z_{i2} \tag{15}$$

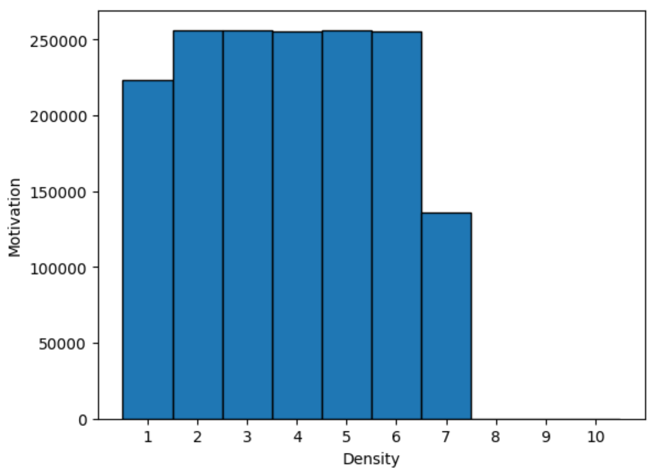

Figure 9. Histogram yielded through the Monte Carlo method on the two-dimensional input space. This shows the frequency of each motivation level based on the reduced input space.

Figure 10 shows that the second principal component, z_{i2}, is mostly negative, despite its corresponding weight also being negative (see Equation (15)). Moreover, this principal component does not appear to be a latent factor influencing student motivation, whereas the first principal component does. This is evident from the figure, as the vectors on the left side correspond to the highest motivation levels, while those on the opposite side are associated with the lowest levels.

In Figure 10, the vectors representing the students who participated in the survey are shown, classified according to the function g obtained from the regression model, as illustrated by the contour lines. Notably, these lines indicate the absence of vectors in the highest motivation level, which explains why this level is unlikely in the simulation outcomes.

VII. CONCLUSION AND FUTURE WORK

In this study, we found that engineering students enrolled in scientific computing courses at the University of Córdoba exhibit a moderate level of motivation, which is the most likely outcome. This suggests that these students find it challenging to grasp the concepts, foundations, and methods taught in these courses. As such, understanding the underlying factors that contribute to motivation is critical in designing more effective learning environments.

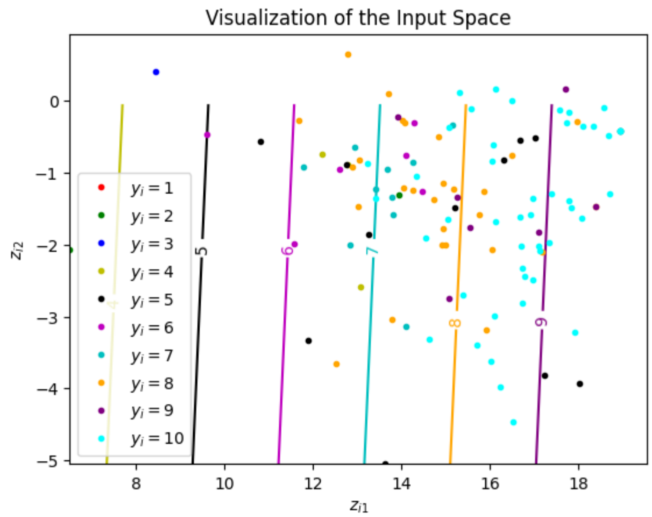

Figure 10. Visualization of the latent factors derived from the regression model. The contour lines show the lower probability of obtaining the higher motivation levels.

As a consequence, it is essential for lecturers to develop effective motivation strategies tailored to the unique challenges of scientific computing courses. Strategies might include more interactive teaching methods, real-world problem-solving tasks, or personalized feedback systems aimed at increasing student engagement. By aligning teaching practices with the motivational needs of students, there is a greater chance of improving learning outcomes and overall academic success.

However, the findings are subject to some limitations. The data collected in this study might not fully capture the diversity of motivation levels across different institutions or disciplines. Therefore, further data collection is necessary to address potential validity threats, such as the potential for selection bias or the lack of data from different academic backgrounds. Expanding the dataset to include students from other universities or fields of study could provide a more comprehensive understanding of the factors influencing student motivation.

So far, to interpret the prediction function, we have assumed a linear relationship between the factors influencing motivation and the student's motivation level. While this assumption has provided valuable insights, it might not fully reflect the complexity of motivation. For future research, we will explore nonlinear regression models such as Gaussian processes, Kernel

Ridge Regression, and Random Forests, which might uncover more nuanced relationships between the variables.

Additionally, we will also evaluate alternative models for dimensionality reduction. Methods, such as non-negative matrix factorization, autoencoders, and t-SNE might be adopted to better capture the underlying structure of the data. These methods may offer advantages over principal component analysis, particularly in terms of capturing nonlinearities or latent factors that influence motivation.

The impact of further data collection should not be underestimated. By collecting more data, we could improve the robustness of our findings and potentially develop a model that can predict motivation levels more accurately across diverse student populations. Future work could also involve incorporating more detailed demographic information, which could lead to insights into how motivation varies across different student groups based on age, prior experience, or other factors.

Finally, future work will include practical steps to further investigate these findings. For example, we plan to collaborate with faculty members in scientific computing courses to apply the insights gained from this study in real-world teaching settings. Pilot studies may also be conducted to test the efficacy of the proposed motivational strategies and validate the results through student feedback and performance.

In summary, while this study provides a valuable starting point, there is much more to explore regarding the complexities of student motivation in scientific computing. We hope that future research will contribute to the development of more effective and personalized teaching methods that foster greater student engagement and success.

ACKNOWLEDGMENT

Caicedo-Castro thanks the Lord Jesus Christ for blessing this project. The authors thank Universidad de Córdoba in Colombia for supporting this study. They also thanks all students who collaborated by answering the survey conducted for collecting the dataset used in this study. Finally, the author thanks the anonymous reviewers for their comments that contributed to improve the quality of this article.

REFERENCES

[1] I. Caicedo-Castro, M. Macea-Anaya, and S. Rivera-Castaño, "Early Forecasting of At-Risk Students of Failing or Dropping Out of a Bachelor's Course Given Their Academic History - The Case Study of Numerical Methods," in *PATTERNS 2023: The Fifteenth International Conference on Pervasive Patterns and Applications*, ser. International Conferences on Pervasive Patterns and Applications. IARIA: International Academy, Research, and Industry Association, 2023, pp. 40–51.

[4] ——, "An Empirical Study of Machine Learning for Course Failure Prediction: A Case Study in Numerical Methods," *International Journal on Advances in Intelligent Systems*, vol. 17, no. 1 and 2, pp. 25–37, 2024.

[2] I. Caicedo-Castro, "Course Prophet: A System for Predicting Course Failures with Machine Learning: A Numerical Methods Case Study," *Sustainability*, vol. 15, no. 18, 2023, 13950.

[3] ——, "Quantum Course Prophet: Quantum Machine Learning for Predicting Course Failures: A Case Study on Numerical Methods," in *Learning and Collaboration Technologies*, P. Zaphiris and A. Ioannou, Eds. Cham: Springer Nature Switzerland, 2024, pp. 220–240.

[5] L. Ayebale, G. Habaasa, and S. Tweheyo, "Factors Affecting Students' Achievement in Mathematics in Secondary Schools in Developing countries: A Rapid Systematic Review," *Statistical Journal of the IAOS*, vol. 36, pp. 1–4, 2020.

[6] M. Gómez-García, H. Hossein-Mohand, J. M. Trujillo-Torres, H. Hossein-Mohand, and I. Aznar-Díaz, "Technological Factors That Influence the Mathematics Performance of Secondary School Students," *Mathematics*, vol. 8, no. 11, 2020, 1935.

[7] J.-M. Trujillo-Torres, H. Hossein-Mohand, M. Gómez-García, H. Hossein-Mohand, and F.-J. Hinojo-Lucena, "Estimating the Academic Performance of Secondary Education Mathematics Students: A Gain Lift Predictive Model," *Mathematics*, vol. 8, no. 12, 2020, 2101.

[8] M. Maamin, S. M. Maat, and Z. H. Iksan, "The Influence of Student Engagement on Mathematical Achievement among Secondary School Students," *Mathematics*, vol. 10, no. 1, 2022, 41.

[9] A. Brezavšček, J. Jerebic, G. Rus, and A. Žnidaršič, "Factors Influencing Mathematics Achievement of University Students of Social Sciences," *Mathematics*, vol. 8, no. 12, 2020, 2134.

[10] E. Martinez-Villarraga, I. Lopez-Cobo, D. Becerra-Alonso, and F. Fernández-Navarro, "Characterizing Mathematics Learning in Colombian Higher Distance Education," *Mathematics*, vol. 9, no. 15, 2021, 1740.

[11] J. Park, S. Kim, and B. Jang, "Analysis of Psychological Factors Influencing Mathematical Achievement and Machine Learning Classification," *Mathematics*, vol. 11, no. 15, 2023, 3380.

[12] S. Batista-Toledo and D. Gavilan, "Student Experience, Satisfaction and Commitment in Blended Learning: A Structural Equation Modelling Approach," *Mathematics*, vol. 11, no. 3, 2023, 749.

[13] M. Charalambides, R. Panaoura, E. Tsolaki, and S. Pericleous, "First Year Engineering Students' Difficulties with Math Courses- What Is the Starting Point for Academic Teachers?" *Education Sciences*, vol. 13, no. 8, 2023, 835.

[14] T. T. Wijaya, B. Yu, F. Xu, Z. Yuan, and M. Mailizar, "Analysis of factors affecting academic performance of mathematics education doctoral students: A structural equation modeling approach," *International Journal of Environmental Research and Public Health*, vol. 20, no. 5, 2023, 4518.

[15] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[16] N. Metropolis and S. Ulam, "The Monte Carlo Method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949.